Recognition of Hand Signs Based on Geometrical Features using Machine Learning and Deep Learning Approaches

Kulandai Josephine Julina Joseph[a], Sree Sharmila Thangaswamy[b]

Abstract
Hand gestures are used by specially challenged people for interacting with the outside world. The gesture vocabulary is created with action words and alphabets with a focus on using them as a sign language interpreter. This paper focuses on the usage of the American Sign Language (ASL) gestures and recognition of few action gestures using single hand. The skin region is identified by marking the skin-colour pixels in Hue Saturation and Value (HSV) colour space. The hand contours are obtained after performing background subtraction techniques and morphological operations. The pre-processed hand mask is given as input to the Support Vector Machine (SVM), a machine learning classifier model, and to the Convolution Neural Network (CNN), a deep learning model. The performance analysis in the predicted output for all the models are analyzed. The accuracy of 97% is witnessed for recognition of ASL gestures and 93% for action gestures using CNN approach that surpasses SVM approach yielding an accuracy of 84% and 80% for the same.
Keywords: American Sign Language; Centroid; Contours; Convolutional Neural Network; Gesture recognition; Hand detection; Support Vector Machine

1. Introduction

Gestures can be used for communication and for controlling purposes (Trigueiros et al., 2012) and it carries some meaningful information. Hand signs are very much useful for communication as it reduces the gap between specially challenged and normal persons. The traditional Machine Learning (ML) models are slower in making predictions. The Deep Learning (DL) model like Convolution Neural Network (CNN) outperformed ML models realizing faster recognition rate (Sreedhar et al., 2022). The motivation behind carrying out the work of hand gesture recognition (HGR) is to create a sign language interpreter for hearing impaired individuals.

Gestures can be captured using a simple digital camera, webcam integrated with the system, depth-based cameras such as Microsoft Kinect, infrared cameras to handle varying illumination, ToF cameras (Time of Flight) that has depth sensors to handle 3D data, and cameras in mobile devices (Berman & Stern, 2012; Z.-h. Chen et al., 2014; Kılıboz & Güdükbay, 2015; Raheja et al., 2015; Su et al., 2020). Gestures are formed using either single hand or double hand (Pradhan et al., 2012) and can be categorized as static or dynamic based on the restrictions of hand movements (Ben Jmaa et al., 2016; Geetha et al., 2013). The static gestures can mostly be recognized using neural network-based approach (Kılıboz & Güdükbay, 2015). Dynamic gestures are an ordered sequence of directional movements and can be recognized mostly using techniques like Finite State Machine (FSM) and Hidden Markov Models (HMM) (Kılıboz & Güdükbay, 2015). Other techniques include Kalman filtering, soft computing methods like Fuzzy logic (Mitra & Acharya, 2007; Yuvaraj & Ragupathy, 2022), traditional machine learning approaches like Support Vector Machine (SVM), and deep learning approaches using Convolutional Neural Network (CNN). The hand trajectory information can be converted to representation format using Needleman-Wunsch sequence matching algorithm (Kılıboz & Güdükbay, 2015).

The detection of hand can be easily achieved using gloves, wearable markers and many such similar devices to track the position of hand to get hand co-ordinates (Kılıboz & Güdükbay, 2015). The gestures can also be drawn using special devices like stylus (Jayamini & Withanage, 2015). However, the vision-based processing of hand regions is cheaper than glove-based or marker-based approach (Ben Jmaa et

[a,b] Department of Information Technology, Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam – 603110, Tamil Nadu, India.
E-mail: [a]josephinejulinajk@ssn.edu.in, [b]sreesharmilat@ssn.edu.in

al., 2016; Geetha et al., 2013). Some of the well-known models in detecting the presence of hand in a raw gesture image is by analyzing shape and appearance of hand. The deformable model uses point-distribution method to analyze the presence of hand (Ben Jmaa et al., 2016). The appearance-based models use Gabor wavelets. There is a possibility of increased complexity when handling more degrees of freedom as in the case of tracking dynamic trajectory, accessing large databases and processing depth information (Dardas & Petriu, 2011; Kılıboz & Güdükbay, 2015; Singha et al., 2018; Trigueiros et al., 2012). It is still more challenging when it involves 3D dynamic signs (Geetha et al., 2013).

In this paper, finger spelling approach denoting the usage of single-handed American Sign Language (ASL) gestures and action gestures are considered. It follows a tactile approach which is commonly used by a person with hearing and visual impairments. The signer is a person who makes a meaningful gesture in front of a camera. The hand gestures made by a signer are captured using the webcam integrated with the system. These are trained by classifier models and recognition results are obtained. Some of the research highlights are as follows. Gesture vocabulary dealing with American Sign Language (ASL) hand signs are created for the study. Pre-processing is carried out resulting in efficient extraction of hand regions from the input feed. Geometrical features of hand regions are given to the model classifiers namely Support Vector Machine and Convolution Neural Network.

The remaining section of the paper is organized as follows: Section 2 provides details about related work carried out in hand detection and throws light on the state-of-the-art methods involved in pre-processing, feature extraction and recognition techniques; Section 3 highlights the significance of the proposed method in detecting hand contours using geometrical features and vision-based processing techniques; Section 4 elaborates on learning aspects in predicting the sign of the hand gesture and showcases the recognition results of ASL and action gestures with emphasize on the qualitative aspects involved in performance evaluation; Section 5 incorporates concluding remarks and future work followed by references.

2. Literature survey

2.1 Pre-processing

Pre-processing is an essential step in detecting the Region of Interest (ROI) in an image. It helps in removal of noise artifacts. Resize, contrast adjustment, histogram equalization and image normalization proved to be good candidates for pre-processing of images (Nachamai, 2013). The list of few prominent pre-processing techniques available in the literature is given in Table 1.

Table 1. **Pre-processing techniques**

Pre-processing techniques	Inferences
Bilateral filtering	Applied on localized hand (Ben Jmaa et al., 2016)
Median filtering	Removes salt and pepper noise (Ben Jmaa et al., 2016)
Edge closing	Simple low-level processing (Ben Jmaa et al., 2016)
Hand filling	Applied after edges are closed to obtain binary image (Ben Jmaa et al., 2016)
Morphological operations	Erosion removes small sharp unwanted details(Lahiani et al., 2016), Dilation for smoothening (Dardas & Petriu, 2011)
Component based sliding window technique	Smoothening and noise removal (Geetha et al., 2013)
Sub-pixel localization	Scale and location variance eliminated (Ben Jmaa et al., 2016)

2.2 Colour spaces

The conventional way of detecting hand in an image lies in identifying skin colour regions in colour space. In the conventional method dominant blob region is labeled as the face and second biggest as the hand. Some of the colour spaces for segmenting skin regions are listed in Table 2.

Table 2. **Colour spaces**

Colour spaces	Inferences
RGB	Simple but do not linearly correspond to human perception and highly correlated red (R), blue (B) and green (G) component (Elgammal et al., 2009); Conversion to gray scale eliminates luminance (Jayamini & Withanage, 2015)
HSI, HSV / HSB, HSL	Robust against rotation, scaling & lighting (Ahuja & Singh, 2015); Perceptual colour spaces and separates hue (H), saturation (S) and the brightness (I, V or L) (Ghotkar et al., 2012)
YUV, YIQ, YC_bC_r	Y represents illumination channel and UV, IQ and CbCr represents orthogonal chrominance channels. TV colour spaces used in image and video compression, separates luminance and chrominance components, and can be detected easily (Dardas & Petriu, 2011; Elgammal et al., 2009; Shin et al., 2006)

2.3 Feature extraction

The shape is an important feature in understanding the ROI in an image (Ghotkar et al., 2012). The extraction of hand contours helps in the discovery of hand region in an image. The convex hull covers the hand enclosure, and it detects the region boundary (Dardas & Petriu, 2011; Farooq & Ali, 2014; Lahiani et al., 2016). This helps in fingertip identification by eliminating false vertices and skeletonization is done to get the accurate tip positions (Nguyen et al., 2015). The shape of the hand is obtained by examining various parameters as given in the Figure 1.

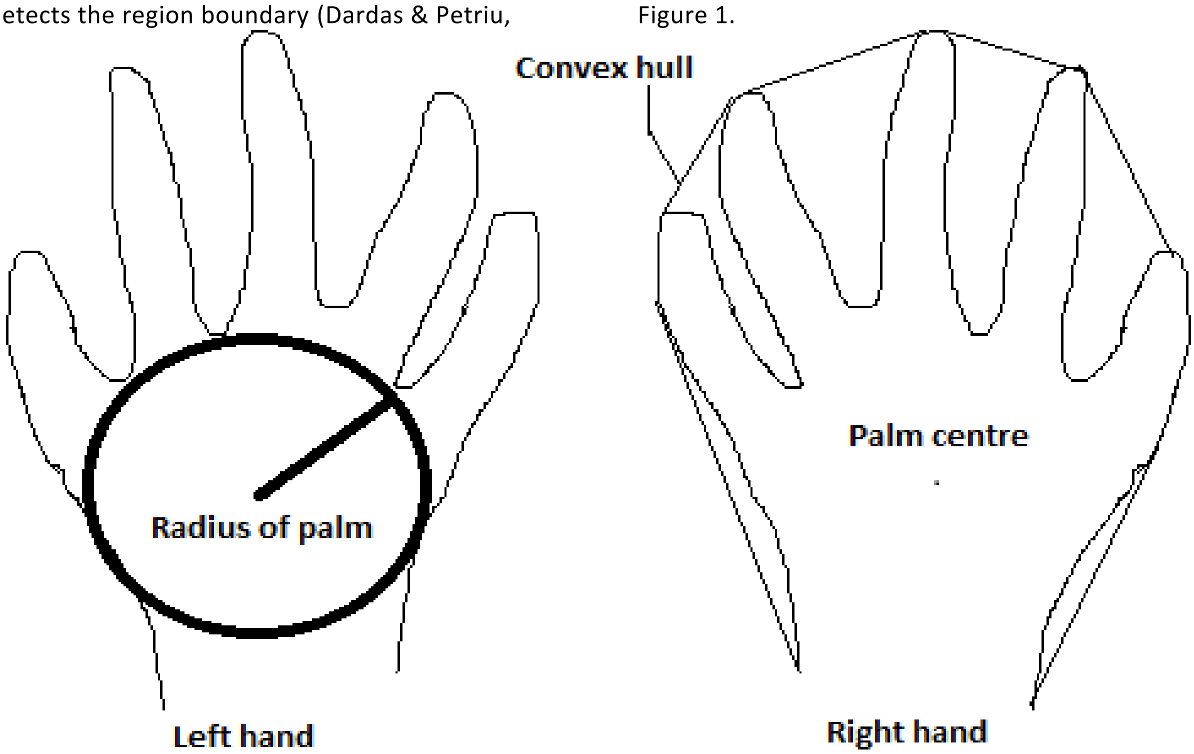


Figure 1. **Hand Feature Candidates**

Hand dimensions and orientation of the hand is measured by studying hand shape fitting in elliptical boundary (Ben Jmaa et al., 2016). The 3D feature descriptors use global information like movement of the centroid and varying depth value (Ben Jmaa et al., 2016). Feature extraction based on new operators is helpful in recognition of hand gestures in real-time (Feng et al., 2011). Some of the feature extraction techniques are listed in Table 3.

Table 3. **Feature extraction techniques**

Feature extraction techniques	**Inferences**
Haar features	Key point localization (Dardas & Petriu, 2011; Feng et al., 2011); Adaboost (Mathias & Turk, 2004)
Edge	Contour identification (Badi et al., 2015)
Geometric features	Mean and variance of the gray pixels, blob area, perimeter and number of convexity defects, hand orientation, histogram, radial signature (Trigueiros et al., 2012)
Speeded-Up Robust Feature (SURF)	Detects rotation; fails scaling (Ghotkar et al., 2012)
Scale Invariant Feature Transform (SIFT)	Illumination and rotation invariant (Feng et al., 2011; Geetha et al., 2013; Jayamini & Withanage, 2015; Nachamai, 2013)
Histogram of Oriented Gradients (HOG)	Translation variance (Pradhan et al., 2012; Trigueiros et al., 2012; Žemgulys et al., 2018), edge orientation preserved (J. K. J. Julina & Sharmila, 2017; J. K. Josephine Julina & Sharmila, 2019)
Hand contours	Hand postures identified contours comparison algorithm (Badi, 2016; Dardas & Petriu, 2011; Mei et al., 2015)

2.4 Recognition

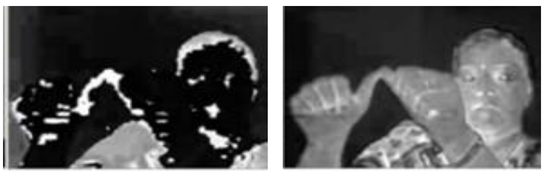
Recognition follows vision based, model based or state-based approaches (F.-S. Chen et al., 2003). Distance metrics like Hausdorff distance calculations and finding Euclidean distance in the feature space (Mathias & Turk, 2004; Mei et al., 2015) is prominently used in calculating hand geometry. Eigen distance is used to calculate the key points distance and it helps in measuring similarity or dissimilarity (Bakina & Mestetskiy, 2011; Geetha et al., 2013; Pradhan et al., 2012; Sreedhar et al., 2022). The recognition of hand gestures and its role in classifying them is given in Table 4.

Table 4. **Recognition techniques**

Recognition techniques	Inferences
Principal Component Analysis	Statistical modeling and images are trained using Eigen vectors and test weights are calculated and is computationally intensive (Ahuja & Singh, 2015; Bakina & Mestetskiy, 2011; Dardas & Petriu, 2011; Nguyen et al., 2015; Pradhan et al., 2012)
Support Vector Machine	Statistical learning theory works on high dimensional data (J. K. J. Julina & Sharmila, 2017; Trigueiros et al., 2012)
Genetic algorithm	Feature selection is efficient (Kılıboz & Güdükbay, 2015)
Naïve Bayes	Probabilistic approach (Trigueiros et al., 2012)
Neural network	Accurate recognition, but cost is high (Vivek & Swaminathan, 2013). Mathematical computational model (Trigueiros et al., 2012)
Hidden Markov Model	Double stochastic process governed by Markov chain and is best suited for dynamic recognition (F.-S. Chen et al., 2003; Mitra & Acharya, 2007)
Inertial motion capture algorithms	Dynamic hand movements recognized (Kılıboz & Güdükbay, 2015)
Self-Organizing Map	Computationally intensive (Trigueiros et al., 2012)
B-spline	Trajectory information and curvature of key points tracked (Geetha et al., 2013)
Random Forest	Joint angles are identified in multi-layer random forest (Ben Jmaa et al., 2016)

3. Methodology

The American Sign Language (ASL) is represented using single hand. The system is built to perform feature learning and classification. The steps in hand gesture recognition include capturing gesture input, applying pre-processing to segment the hand region and extracting relevant features to carry out recognition tasks. The hand signs are captured and the input video showing different gestures of hand is stored. The custom dataset is created in real time. The captured scene from the input video feed is converted into frames and the noise in the frame is removed by applying pre-processing techniques. The RGB colour frames are converted to gray scale to reduce the computational cost. The noise factors degrade the performance of the system and hence utmost care must be taken in pre-processing the images. The output of a skin segmentation is shown in Figure 2 for HSV colour space.



a. b.

Figure 2. **Colour space a. Hue b. Saturation**

The region to be segmented is purely based on the type of application such as recognition of facial biometrics or hand geometry extraction. The face region is removed by localizing the facial occurrence in the frame based on the Haar features (Viola & Jones, 2004). The removal of face is carried out by subtracting face regions by means of template matching approach as shown in Figure 3.

Figure 3. **Face template match**

The algorithm 1 specifies the steps involved in the face region identification. The hand mask is created by considering the morphological operations and the steps are mentioned in algorithm 2. The entire pre-processing workflow is given in the Figure 4.

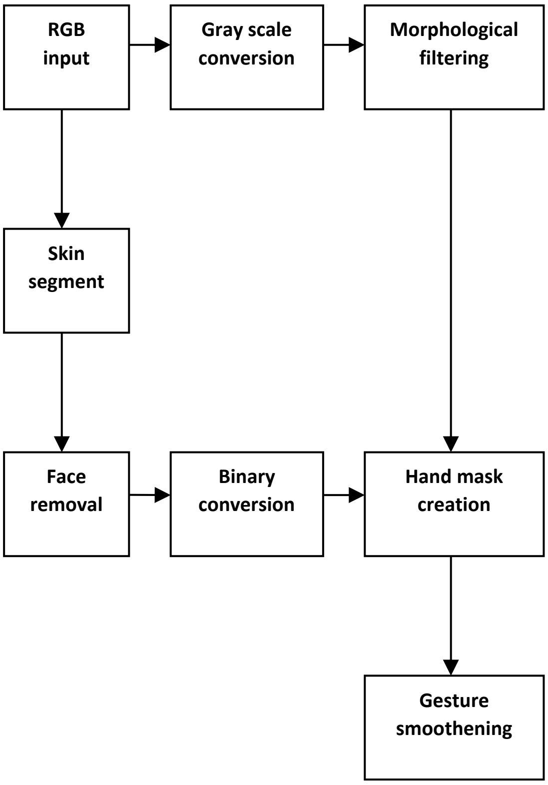


Figure 4. **Preprocessing workflow**

Certain gestures appear to be different when left and right hands are used separately as one is the mirror image of the other. The image is prepared for recognition after it undergoes pre-processing techniques and is shown in the Figure 5. The hand mask is created after skin segmentation.

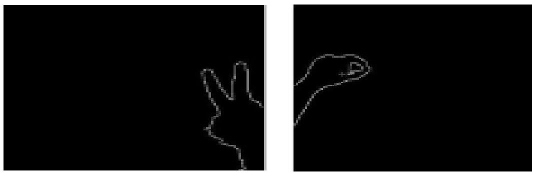
Input: Input frame, face template (4 box co-ordinates of face) and template size
Output: Marked Face Image
Step1: correlationOp ← Compute normalized 2-D cross-correlation of the face template and input image
Step2: correlationVal ← Find the maximum correlation value of correlationOp
Step3: [y, x] ← Mark equivalent rows and columns in the correlationOp based on correlationVal
Step4: Return the boxed offset
○ Offset ← (x-templatesize) (y-templatesize)

Algorithm 1. **Face template match**

Figure 5. **Hand Mask**

The input to the recognition model is the captured image after the background subtraction process is completed. It involves preparation of the foreground mask by applying bitwise and of the kernel and the input frame. The segmented hand region is given as input to the recognition models namely SVM and CNN.

The basic features comprise of appearance, shape, colour and contrast (Viola & Jones, 2004). Image contrast represents local orientation of hand and is denoted by horizontal and vertical image pixel differences (Trigueiros et al., 2012). Some of the sample hand contours are shown in Figure 6.



(a) (b)

Figure 6. (**a) Gesture V (b) Gesture O**

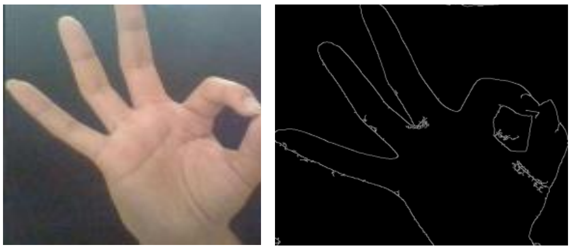
Input: Image frame
Output: Hand mask
Step1: fgmask ← Capture the background model and subtract it from the captured input gesture frame based on the background model learning rate of video feed
Step2: Initialize a kernel of array of ones for performing morphological transformations
Step3: Segment the foreground hand gesture by applying the kernel
○ fgmask ← erode (fgmask)
Step4: Mask ← bitwise_and (fgmask, kernel)

Algorithm 2. **Hand mask creation**

A set of images are trained with SVM and CNN models. By using these models, the output is predicted for the given input gesture. The trained model will label the images based on the hand signs. The system developed also recognizes few action gestures. The background model is captured, and sample image in action dataset and the tracing of contours is shown in the Figure 7.

Input: Segmented hand region G (x, y)
Output: Feature Points
Step1: Trace the contours by finding perimeter pixels of the objects
Step2: Fix the centroid and finger co-ordinates
Step3: Obtain the tips of fingers and get the convex hull region
Step4: Store the marked feature points for use in recognition

Algorithm 3. **Feature map creation**



(a) (b)

Figure 7. **(a) Input (b) Contour**

The hand region is focused and the geometric feature points marking especially centroid are carried out which is depicted in the Figure 8.

(a) (b)

Figure 8. **(a) Contour (b) Centroid**

Some of the limitations seen in this experiment is inability to handle challenges that arises due to varying illumination conditions. Otherwise, the study paved the way for successful recognition of almost all hand signs of ASL gestures and few normal action gestures. The future research direction involves conducting the study on individuals who suffer from hearing loss creating a sign language interpreter.

4. Results

A hand gesture recognition system is developed for American Sign Language (ASL) to recognize finger spelling using a single hand. The learning algorithms used in this work consists of traditional ML method namely SVM and DL method namely CNN models to carry out the recognition tasks. The pre-processed data is splitted into training and testing sets. The dataset for 26 letters of English alphabets in ASL representation and 5 action gestures namely “Hello, Okay, Good, Bad and Peace” are created. Hyperparameters optimization are carried out and the feature selection is given to both cross validation model and trained model. The over fitting problems in SVM is not seen in CNN.CNN is more suitable for improving accuracy of image classification tasks. CNN is a multilayered feed forward neural network mostly used in image recognition. The sequential model is built with four convolutional layers and two fully controlled layers as depicted in the Figure 9. The Convolutional layers namely Conv2D layer, batch normalization, rectified linear activation function and max pooling 2D layer are used for feature extraction. The flattened output is given to fully connected layers. The fully connected layers are used for interpreting the gestures. It includes dense layer, batch normalization, rectified linear activation function and drop out layer. Then the soft max functions are applied. To reduce the over fitting problems, different techniques including dropout and batch normalization is used.

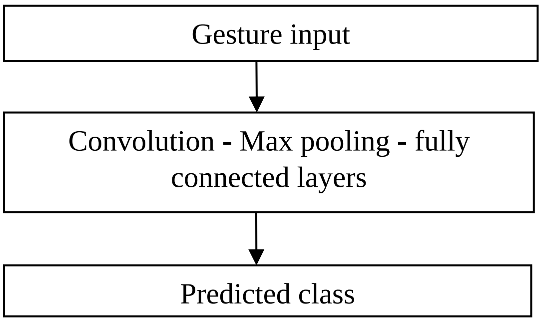


Figure 9. **CNN model**

The model is evaluated based on accuracy and the weights associated with the higher validation accuracy is saved. The model loss is calculated, and it is almost zero if the prediction is perfect and is greater otherwise. The feature maps are created from the Convolutional layers and fully connected layers to produce the output of the gesture classes.

The performance of the classification model is evaluated by means of accuracy, sensitivity, and specificity. Accuracy is calculated using the Eq. (1). It is the ratio of the number of correct predictions to total number of correct and incorrect predictions.

$$\text{Accuracy}=\frac{TP+TN}{(TP+TN+FP+FN)} \quad (1)$$

TP refers to True Positives, TN refers to True Negatives, FP refers to False Positives, FN refers to False Negatives. The specificity and sensitivity are given in Eq. (2) and Eq. (3) and is calculated based on correct and incorrect predictions.

Specificity= $\frac{TN}{(TN+FP)}$ (2)

Sensitivity= $\frac{TP}{(TP+FN)}$ (3)

The gesture recognition results are given in Figure 10 which detects ASL gestures and a set of action words.

BAD

(a) (b)

Figure 10. **(a) Input (b) Recognition**

The performance of the model using SVM, and CNN approach is given in the Table 5.

Table 5. **Performance metrics**

Classifier	Type	Performance measures	Results (%)
SVM	ASL	Accuracy	84.78
		Specificity	92.13
		Sensitivity	94.47
CNN		Accuracy	97.12
		Specificity	98.53
		Sensitivity	96.25
SVM	Action	Accuracy	80.78
		Specificity	78.53
		Sensitivity	76.79
CNN		Accuracy	93.33
		Specificity	94.51
		Sensitivity	91.45

The accuracy of 97% is witnessed for recognition of ASL gestures and 93% for action gestures using CNN approach that surpasses SVM approach yielding an accuracy of 84% and 80% for the same.

5. Conclusion

Hand gesture detection and recognition is a complex task and is an active research field as it finds its use in many applications namely gaming, robotics, traffic control, mobile devices, video surveillance, sign language interpretation, virtual and augmented reality, expression analysis, human computer and human robot interaction and many other applications of the same kind. Smart phones have been employed with gesture controlling features. In this paper, study of hand gestures using ASL representation and few action gestures is considered and no doubt it can be used as a sign language interpreter. The result showed the CNN model outperform SVM approach. The future work involves capturing dynamic signs using depth sensors and motion controllers namely Microsoft Kinect and Leap Motion Controller.

References

[1] Ahuja, M. K., & Singh, A. (2015). Static vision based Hand Gesture recognition using principal component analysis. *2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE)* (pp. 402-406). IEEE. https://doi.org/10.1109/MITE.2015.7375353

[2] Badi, H. (2016). Recent methods in vision-based hand gesture recognition. *International Journal of Data Science and Analytics, 1*(2), 77-87. https://doi.org/10.1007/s41060-016-0008-z

[3] Badi, H., Kareem, S. A., & Husein, S. (2015). Feature extraction technique for static hand gesture recognition. *Chaudhary A. Recent Trends in Hand Gesture Recognition, 3*, 19-41. https://sciencegatepub.com/books/gcsr/gcsr_vol3/GCSR_Vol3_Ch2.pdf

[4] Bakina, I., & Mestetskiy, L. (2011). Hand Shape Recognition from Natural Hand Position. *2011 International Conference on Hand-Based Biometrics* (pp. 1-6). IEEE. https://doi.org/10.1109/ICHB.2011.6094317

[5] Ben Jmaa, A., Mahdi, W., Ben Jemaa, Y., & Ben Hamadou, A. (2016). A new approach for hand gestures recognition based on depth map captured by rgb-d camera. *Computación y Sistemas, 20*(4), 709-721. https://doi.org/10.13053/cys-20-4-2390

[6] Berman, S., & Stern, H. (2012). Sensors for Gesture Recognition Systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42*(3), 277-290. https://doi.org/10.1109/TSMCC.2011.2161077

[7] Chen, F.-S., Fu, C.-M., & Huang, C.-L. (2003). Hand gesture recognition using a real-time tracking method and hidden Markov models. *Image and vision computing, 21*(8), 745-758. https://doi.org/10.1016/S0262-8856(03)00070-2

[8] Chen, Z.-h., Kim, J.-T., Liang, J., Zhang, J., & Yuan, Y.-B. (2014). Real-time hand gesture recognition using finger segmentation. *The Scientific World Journal, 2014*, 267872. https://doi.org/10.1155/2014/267872

[9] Dardas, N. H., & Petriu, E. M. (2011). Hand gesture detection and recognition using principal component analysis. *2011 IEEE*

International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA) Proceedings (pp. 1-6). IEEE. https://doi.org/10.1109/CIMSA.2011.6059935

[10] Elgammal, A., Muang, C., & Hu, D. (2009). Skin detection-a short tutorial. *Encyclopedia of Biometrics, 4*, 1218-1224. https://people.cs.rutgers.edu/~elgammal/pub/skin.pdf

[11] Farooq, J., & Ali, M. B. (2014). Real time hand gesture recognition for computer interaction. *2014 International Conference on Robotics and Emerging Allied Technologies in Engineering (iCREATE)* (pp. 73-77). IEEE. https://doi.org/10.1109/iCREATE.2014.6828342

[12] Feng, Z., Yang, B., Chen, Y., Zheng, Y., Xu, T., Li, Y., . . . Zhu, D. (2011). Features extraction from hand images based on new detection operators. *Pattern Recognition, 44*(5), 1089-1105. https://doi.org/10.1016/j.patcog.2010.08.007

[13] Geetha, M., Manjusha, C., Unnikrishnan, P., & Harikrishnan, R. (2013). A vision based dynamic gesture recognition of Indian Sign Language on Kinect based depth images. *2013 International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications (C2SPCA)* (pp. 1-7). IEEE. https://doi.org/10.1109/C2SPCA.2013.6749448

[14] Ghotkar, A. S., Khatal, R., Khupase, S., Asati, S., & Hadap, M. (2012). Hand gesture recognition for Indian Sign Language. *2012 International Conference on Computer Communication and Informatics* (pp. 1-4). IEEE. https://doi.org/10.1109/ICCCI.2012.6158807

[15] Jayamini, U. P. K., & Withanage, D. K. (2015). A Novel Approach in Hand Gesture Recognition for Sign Language. *International Journal of Engineering Research & Technology (IJERT), 4*(3), 219-224. http://dx.doi.org/10.17577/IJERTV4IS030366

[16] Julina, J. K. J., & Sharmila, T. S. (2017). Facial recognition using histogram of gradients and support vector machines. *2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICCCSP.2017.7944082

[17] Julina, J. K. J., & Sharmila, T. S. (2019). Facial Emotion Recognition in Videos using HOG and LBP. *2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)* (pp. 56-60). IEEE. https://doi.org/10.1109/RTEICT46194.2019.9016766

[18] Kılıboz, N. Ç., & Güdükbay, U. (2015). A hand gesture recognition technique for human–computer interaction. *Journal of Visual Communication and Image Representation, 28*, 97-104. https://doi.org/10.1016/j.jvcir.2015.01.015

[19] Lahiani, H., Kherallah, M., & Neji, M. (2016). Hand pose estimation system based on Viola-Jones algorithm for Android devices. *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)* (pp. 1-6). IEEE. https://doi.org/10.1109/AICCSA.2016.7945717

[20] Mathias, K., & Turk, M. (2004). Robust hand detection. *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.* (pp. 614-614). IEEE Computer Society. https://doi.ieeecomputersociety.org/10.1109/AFGR.2004.1301601

[21] Mei, K., Zhang, J., Li, G., Xi, B., Zheng, N., & Fan, J. (2015). Training more discriminative multi-class classifiers for hand detection. *Pattern Recognition, 48*(3), 785-797. https://doi.org/10.1016/j.patcog.2014.09.001

[22] Mitra, S., & Acharya, T. (2007). Gesture Recognition: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 37*(3), 311-324. https://doi.org/10.1109/TSMCC.2007.893280

[23] Nachamai, M. (2013). Alphabet recognition of american sign language: a hand gesture recognition approach using sift algorithm. *International Journal of Artificial Intelligence & Applications, 4*(1), 105-115. https://doi.org/10.5121/ijaia.2013.4108

[24] Nguyen, T.-N., Huynh, H.-H., & Meunier, J. (2015). Static hand gesture recognition using principal component analysis combined with artificial neural network. *Journal of Automation and Control Engineering, 3*(1), 40-45. http://dx.doi.org/10.12720/joace.3.1.40-45

[25] Pradhan, A., Ghose, M., Pradhan, M., Qazi, S., Moors, T., EL-Arab, I., . . . Memon, A. (2012). A hand gesture recognition using feature extraction. *International Journal of Current Engineering and Technology 2*(4), 323-327. https://inpressco.com/wp-content/uploads/2012/12/Paper1323-3272.pdf

[26] Raheja, J. L., Singhal, A., & Chaudhary, A. (2015). Android based portable hand sign recognition system. *arXiv e-prints*. https://doi.org/10.48550/arXiv.1503.03614

[27] Shin, J.-H., Lee, J.-S., Kil, S.-K., Shen, D.-F., Ryu, J.-G., Lee, E.-H., . . . Hong, S.-H. (2006). Hand region extraction and gesture recognition using entropy analysis. *IJCSNS International Journal of Computer Science and Network Security, 6*(2A), 216-222. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.475.9818&rep=rep1&type=pdf

[28] Singha, J., Roy, A., & Laskar, R. H. (2018). Dynamic hand gesture recognition using vision-based approach for human–computer interaction. *Neural Computing and Applications, 29*(4), 1129-1141. https://doi.org/10.1007/s00521-016-2525-z

[29] Sreedhar, P., Satya, S., & Nandhagopal, N. (2022). Classification Similarity Network Model for Image Fusion Using Resnet50 and GoogLeNet. *Intelligent Automation & Soft Computing, 31*(3), 1331-1344.

[30] Su, H., Ovur, S. E., Zhou, X., Qi, W., Ferrigno, G., & De Momi, E. (2020). Depth vision guided hand gesture recognition using electromyographic signals. *Advanced Robotics, 34*(15), 985-997. https://doi.org/10.1080/01691864.2020.1713886

[31] Trigueiros, P., Ribeiro, F., & Reis, L. P. (2012). A comparison of machine learning algorithms applied to hand gesture recognition. *7th Iberian Conference on Information Systems and Technologies (CISTI 2012)* (pp. 1-6). IEEE. http://hdl.handle.net/1822/25749

[32] Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision, 57*(2), 137-154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb

[33] Vivek, V. J., & Swaminathan, P. (2013). Robust hand gesture recognition algorithm for simple mouse control. *International Journal of Computer and Communication Engineering, 2*(2), 219-221. http://www.ijcce.org/papers/175-T30017.pdf

[34] Yuvaraj, K., & Ragupathy, U. (2022). Hybrid Active Contour Mammographic Mass Segmentation and Classification. *Computer Systems Science and Engineering, 40*(3), 823-834. http://dx.doi.org/10.32604/csse.2022.018837

[35] Žemgulys, J., Raudonis, V., Maskeliūnas, R., & Damaševičius, R. (2018). Recognition of basketball referee signals from videos using Histogram of Oriented Gradients (HOG) and Support Vector Machine (SVM). *Procedia computer science, 130*, 953-960. https://doi.org/10.1016/j.procs.2018.04.095